Spatial extreme quantile estimation using a weighted log-likelihood approach

Julie Carreau, Stéphane Girard

Journal de la Société Française de Statistique
Vol. 152 No. 3 (2011)

Spatial extreme quantile estimation using a weighted log-likelihood approach

Titre: Estimation de quantiles extrêmes spatiaux par la méthode de la log-vraisemblance pondérée

Julie Carreau [1] and Stéphane Girard [2]

Abstract: We propose to estimate spatial extreme quantiles by a weighted log-likelihood approach. It is assumed that the conditional distribution of the variable of interest follows a generalized extreme-value distribution. The associated response surfaces are estimated thanks to the introduction of weights in the log-likelihood. These weights depend on the distance between the point of interest and the observations. The construction of a proper distance relies on the combination of a multidimensional scaling unfolding with a neural network regression. Our approach is illustrated both on simulated and real rainfall datasets.

Résumé : Nous proposons d'estimer des quantiles extrêmes spatiaux par une approche de type log-vraisemblance pondérée. Pour cela, nous supposons que, conditionnellement aux covariables, la variable d'intérêt suit une loi des valeurs extrêmes. Les surfaces de réponse associées sont estimées en introduisant des poids dans la log-vraisemblance. Ces poids dépendent de la distance entre le point d'estimation et les observations. La construction d'une distance appropriée repose sur la combinaison d'une étape de dépliage par "multidimensional scaling" et d'une étape de régression par réseaux de neurones. Notre approche est illustrée sur des jeux de données réelles et simulées.

Keywords: generalized extreme-value distribution, semi-parametric estimation, multidimensional scaling, neural networks
AMS 2000 subject classifications: 62G32, 62G05, 62M30

1. Introduction

An important literature is dedicated to the estimation of extreme quantiles, *i.e.* quantiles of high order α. The most popular estimator was proposed in [29], in the context of heavy-tailed distributions. We also refer to [10] for the general case.

When some covariate x is recorded simultaneously with the quantity of interest Y, the extreme quantile, denoted by $q(\alpha,x)$, thus depends on the covariate x and is referred to as the conditional extreme quantile. In our real data study, we are interested in the estimation of return levels associated with extreme rainfalls as a function of the geographical location. In this case, x is a three-dimensional covariate involving the longitude, latitude and altitude, and $q(\alpha,x)$ is called a spatial extreme quantile.

Parametric models for conditional extremes are proposed in [7, 24] whereas semi-parametric methods are considered in [1, 14]. Fully non-parametric estimators have been first introduced in [8], where a local polynomial modeling of the extreme observations is used. Similarly, spline

1. HydroSciences Montpellier
E-mail: julie.carreau@univ-montp2.fr
2. Team Mistis, INRIA Rhône-Alpes and LJK
E-mail: stephane.girard@inrialpes.fr

 ISSN: 2102-6238

estimators are fitted in [4] through a penalized maximum likelihood method. These results are extended to multidimensional covariates in [2] where local polynomial estimators are proposed. Kernel estimators are proposed in [6, 12] when the conditional distribution of Y conditionally to $X = x$ is heavy-tailed.

We propose here to estimate the spatial extreme quantile by a weighted log-likelihood approach. We assume that, conditionally on $X = x$, Y follows a generalized extreme-value (GEV) distribution. Its parameters at point x, the so-called response surfaces, are estimated thanks to the introduction of weights in the log-likelihood. These weights depend on the distance $d(x,X)$ between the point of interest x and the covariate X. The construction of a proper distance is an important part of this work. As an illustration, in the above mentioned climatology study, the estimation of the return level at a given geographical point is based on rainfalls measured at the neighboring raingauges. The paper is organized as follows: The spatial extreme quantile estimator is introduced in Section 2 and the construction of the distance is detailed. The performances of the estimator are illustrated both on simulated and on real data in Section 3. Finally, our future work is discussed in Section 4.

2. Spatial extreme quantile estimator

Let (X_i,Y_i), $i = 1,\dots,n$ be independent copies of a random pair (X,Y) in $\mathbb{R}^p \times \mathbb{R}$. We assume that, conditionally on $X = x$, Y follows a GEV distribution with density

$$\begin{aligned} f(y;\mu(x),\sigma(x),\xi(x)) &= \frac{1}{\sigma(x)}\left[1+\xi(x)\left(\frac{y-\mu(x)}{\sigma(x)}\right)\right]_+^{-1/\xi(x)-1} \\ &\times \exp\left\{-\left[1+\xi(x)\left(\frac{y-\mu(x)}{\sigma(x)}\right)\right]_+^{-1/\xi(x)}\right\} \end{aligned} \tag{1}$$

where $y \in \mathbb{R}$ and $[\cdot]_+ := \max(\cdot,0)$. Here, the location parameter $\mu(x)$, the scale parameter $\sigma(x)$ and the extreme-value index $\xi(x)$ are unknown functions $\mathbb{R}^p \to \mathbb{R}$ of the covariate x. The GEV distribution is the limit distribution of properly normalized maxima of a sequence of independent and identically distributed random variables, see for instance [11]. This justifies the GEV distribution as an appropriate model for maxima of long sequences of random variables.

The estimation of these response surfaces is addressed using a weighted log-likelihood approach in Subsection 2.1. The definition of weights requires the construction of a proper distance on $\mathbb{R}^p$. This point is discussed in Subsection 2.2.

2.1. Response Surface Estimator

2.1.1. Weighted negative log-likelihood

We propose to use all the observations available to estimate the GEV parameters of a given variable Y, not only the observations sampled from Y. Weights in the negative log-likelihood control the influence of each observation on the estimation. These weights are computed by applying a positive kernel function K with bandwidth parameter h to the distance between covariates. Large values of h permit to pooling observations with similar covariates to construct the weighted log-likelihood. The underlying motivation is that the distribution of the variable of

 ISSN: 2102-6238

interest Y is similar for covariates X close to each other. For instance, in the rainfall application, X represents the geographical coordinates and Y is the rainfall maxima. Since the sample of maxima at a given location X can be small (it can be zero if X corresponds to an ungauged site), estimation based on observations at this location alone ($h = 0$) can be unreliable. Conversely, if $h > 0$ is large enough, rainfalls of neighbor stations are also taken into account. This bias-variance tradeoff is addressed in [18].

For all $x \in \mathbb{R}^p$, we propose to minimize the following weighted negative log-likelihood with respect to (μ, σ, ξ):

$$\ell(\mu,\sigma,\xi) = -\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) \log f(Y_i;\mu,\sigma;\xi)}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)}, \tag{2}$$

where $d(\cdot,\cdot)$ is a distance on $\mathbb{R}^p$, K is the kernel function and h the bandwidth.

The theoretical properties of estimators minimizing the negative weighted log-likelihood are established in [25] in the regular case. These results do not apply here since the support of the GEV density (1) depends on the parameters. Note that the asymptotic properties of the (unweighted) maximum likelihood estimators have been established very recently [31, 32].

In the next paragraph, we focus on the implementation aspects. The asymptotic issues are beyond the scope of this paper.

2.1.2. *Minimization*

First, it is assumed that $\xi(x) > 0$ and the following re-parametrization is considered: Letting

$$(\theta(x),\beta(x),k(x)) := (\mu(x) - \sigma(x)/\xi(x), \xi(x)/\sigma(x), 1/\xi(x)),$$

see also [7], the GEV density (1) can be rewritten as

$$\begin{aligned} \tilde{f}(y;\theta(x),\beta(x),k(x)) &= k(x)\beta(x)^{-k(x)}[y-\theta(x)]_+^{-k(x)-1} \\ &\times \exp\left\{-\beta(x)^{-k(x)}\,[y-\theta(x)]_+^{-k(x)}\right\}. \end{aligned} \tag{3}$$

Replacing in (2), the quantity to minimize is

$$\begin{aligned} \tilde{\ell}(\theta,\beta,k) &= -\log(k) + k\log(\beta) + (k+1)\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)\log[Y_i-\theta]_+}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} \\ &+ \beta^{-k}\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)[Y_i-\theta]_+^{-k}}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)}. \end{aligned} \tag{4}$$

 ISSN: 2102-6238

Taking the derivative of (4) with respect to β and equating to zero, we get the following expression

$$\beta(\theta,k) = \left(\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) [Y_i-\theta]_+^{-k}}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} \right)^{1/k}.$$

Replacing in (4), the function to minimize is

$$\begin{aligned} \check{\ell}(\theta,k) &= -\log(k) + (k+1) \frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) \log[Y_i-\theta]_+}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} \\ &+ \log\left(\frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) [Y_i-\theta]_+^{-k}}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} \right), \end{aligned}$$

or equivalently

$$\begin{aligned} \check{\ell}(\theta,k) &= -\log(k) + (k+1) \frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) \log[Y_i-\theta]_+}{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right)} \\ &+ \log\left(\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{h}\right) [Y_i-\theta]_+^{-k} \right). \end{aligned} \tag{5}$$

Finally, this minimization problem involves only two free parameters k and θ. It can be solved using a grid search algorithm or using dedicated algorithms [17, 20]. The corresponding estimated spatial extreme quantile is given by

$$\hat{q}(\alpha,x) = \hat{\theta}(x) + \frac{1}{\hat{\beta}(x)} [n(1-\alpha)]^{-1/\hat{k}(x)},$$

where $\hat{\theta}(x)$, $\hat{\beta}(x)$ and $\hat{k}(x)$ are the estimated surface responses. We refer to [13] for further details on this derivation and to [21] for the maximum likelihood estimation of GEV parameters in the unconditional case.

2.2. *Distance in Distribution*

A classical way to define a distance $d(\cdot,\cdot)$ which is then used in the kernel function would be to carefully select a set of covariates X. This set of covariates should give information on the extreme rainfall characteristics of raingauges. X could include, in addition to geographical covariates such

 ISSN: 2102-6238

as longitude, latitude and altitude, covariates which describe the topography and atmospheric variables which might be responsible for heavy rainfall [27, 30]. The intuition behind this is that raingauges with similar covariates have similar extreme rainfall distributions which translate into similar GEV parameters according to our assumptions. This distance might be taken as the geographical distance or more generally as the Euclidean distance in the space of the covariates. In the following, we use the notation Z to indicate a raingauge station.

2.2.1. 1D Synthetic Data Study

We first performed a simple one dimensional synthetic data study where $d(\cdot,\cdot)$ is known in order to evaluate empirically the behavior of the GEV parameters estimated at a site with no observations so that observations from neighboring sites only entered the estimation. The setup is the following. There are two sites Z_0 and Z_1 where observations, drawn from a GEV distribution, are available. The distance in distribution is given by $d(Z_0,Z_1)=d_{01}$ and implies the following difference in the GEV parameters:

$$\begin{array}{llll} \mu(X_i)=\mu_0=14, & \sigma(X_i)=\sigma_0=0.7, & \xi(X_i)=\xi_0=0.2 & \text{if } (X_i,Y_i)\in Z_0 \\ \mu(X_i)=\mu_0+d_{01}, & \sigma(X_i)=\sigma_0+{}^{d_{01}}\!/_{20}, & \xi(X_i)=\xi_0 & \text{if } (X_i,Y_i)\in Z_1. \end{array}$$

Here, the notation $(X_i,Y_i)\in Z$ indicates that the couple (X_i,Y_i) has been recorded at site Z. Let Z be an ungauged site between Z_0 and Z_1, that is $d(Z_i,Z)\leq d(Z_0,Z_1)$ with $i\in\{0,1\}$. More precisely, we assume that $d(Z_0,Z)={}^{d_{01}}\!/_{m}$, with $m>1$. To estimate the GEV parameters at site Z, we need to determine the weights of the observations at both sites Z_0 and Z_1. Because the weights are constrained to sum to one, we just need to determine $\lambda\in[0,1]$, the weight of the observations from site Z_1 and thus the observations in the log-likelihood from site Z_0 are weighted by $1-\lambda$. Therefore, the weighted negative log-likelihood (5) can be simplified as:

$$\begin{aligned} \check{\ell}(\theta,k) &= -\log(k)+(k+1)\frac{(1-\lambda)\sum\limits_{(X_i,Y_i)\in Z_0}\log[Y_i-\theta]_+ + \lambda\sum\limits_{(X_i,Y_i)\in Z_1}\log[Y_i-\theta]_+}{(1-\lambda)\sharp((X_i,Y_i)\in Z_0)+\lambda\sharp((X_i,Y_i)\in Z_1)} \\ &+ \log\left((1-\lambda)\sum_{(X_i,Y_i)\in Z_0}[Y_i-\theta]_+^{-k}+\lambda\sum_{(X_i,Y_i)\in Z_1}[Y_i-\theta]_+^{-k}\right) \\ &= -\log(k)+\frac{2(k+1)}{n}\left((1-\lambda)\sum_{(X_i,Y_i)\in Z_0}\log[Y_i-\theta]_+ + \lambda\sum_{(X_i,Y_i)\in Z_1}\log[Y_i-\theta]_+\right) \\ &+ \log\left((1-\lambda)\sum_{(X_i,Y_i)\in Z_0}[Y_i-\theta]_+^{-k}+\lambda\sum_{(X_i,Y_i)\in Z_1}[Y_i-\theta]_+^{-k}\right), \end{aligned} \tag{6}$$

where it has been assumed that $\sharp((X_i,Y_i)\in Z_0)=\sharp((X_i,Y_i)\in Z_1)={}^{n}\!/_{2}$. For a given d_{01} and m, the simulation goes as follows :

1. Generate 100 observations at each site Z_0 and Z_1 according to their GEV distribution.
2. For $\lambda=\{0,0.1,0.2,\ldots,1\}$, estimate the GEV parameters at site Z by minimizing the weighted negative log-likelihood (6).

 ISSN: 2102-6238

3. Evaluate the goodness-of-fit of the estimates by generating observations at site Z according to a GEV with parameters $\mu = \mu_0 + {}^{d_{01}}/_{m}$, $\sigma = \sigma_0 + {}^{d_{01}}/_{20m}$ and $\xi = \xi_0$ and computing the Cramer-Von Mises statistic (see for instance [26]).

The simulation in steps 1-3 is repeated a thousand times. In Figure 1, we present a summary of these simulations for the case in which $d_{01} = 1$ and $m \in \{2,5\}$. In the top left panel of Figure 1 is depicted the median of the Cramer-Von Mises statistic over the replications with respect to the value of λ. In the three panels that follow, in clockwise order, the median and the quantiles of level 0.25 and 0.75 of the GEV parameter estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ are shown as functions of λ.

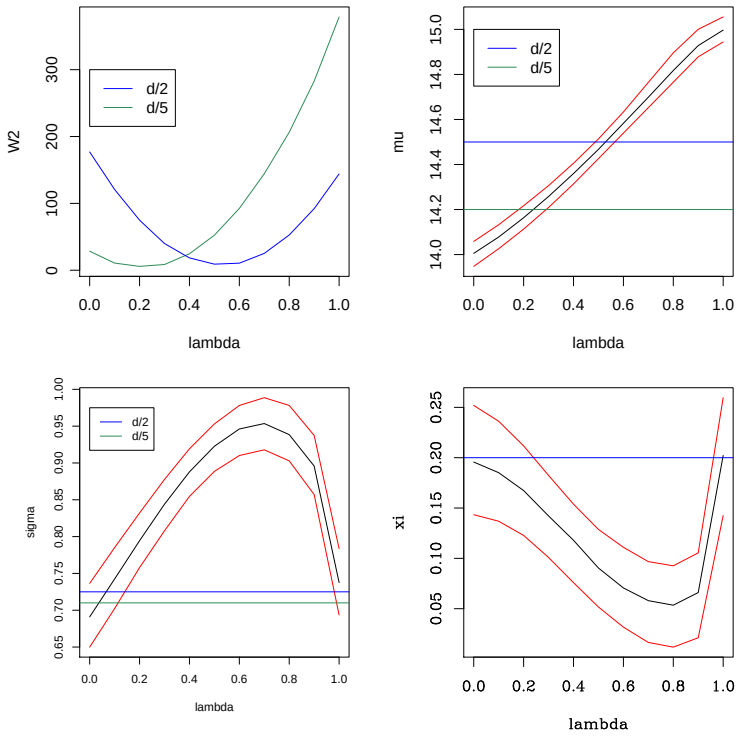


Figure 1: *One dimensional simulation study with a distance in distribution $d_{01} = 1$ between site Z_0 and site Z_1. The GEV parameters are estimated at site $Z = d_{01}/m$, with $m \in \{2,5\}$. In the top left panel is depicted the median Cramer-Von Mises statistic W_2 at site Z over 1000 replications with respect to λ in blue when $m = 2$ and in green when $m = 5$. Then in clockwise order : for each GEV parameter μ, σ and ξ are shown the median and quantiles of level 0.25 and 0.75 of the estimates at site Z over 1000 replications with respect to λ. The blue and green horizontal lines indicate the value of the GEV parameters at site Z for the generative model when $m = 2$ and $m = 5$ respectively (the ξ_0 parameter is equal to 0.2 for $m = 2$ and $m = 5$ so there is just one line).*

The curve of the median of the Cramer-Von Mises statistic as a function of λ indicates a minimum at roughly $\lambda_{\text{opt}}(Z) = {}^1/_2$ for $m = 2$ and $\lambda_{\text{opt}}(Z) = {}^1/_5$ for $m = 5$ where $\lambda_{\text{opt}}(Z)$ is the optimal λ for site Z with respect to the Cramer-Von Mises statistic. As expected, the value of the optimal weights in the log-likelihood estimator is directly related to the distance in distri-

 ISSN: 2102-6238

bution: $\lambda_{\text{opt}}(Z) = d_{01}/m$. The location parameter estimate (top right panel) has no apparent bias when λ is chosen optimally. This is not the case for the scale and tail index estimators (bottom row) where some bias is present. We performed the same simulation with $d_{01} = 2$, that is when the distance in distribution between the sites with observations is twice as large (the results are not shown in this paper). The Cramer - Von Mises statistic indicates the same optimal values for λ. However, the curve is not as smooth as in the case $d_{01} = 1$. In addition, the bias of the GEV parameter estimates is larger in the case $d_{01} = 2$.

From this one dimensional simulation, we conclude that it is possible to get reasonable estimates using the weighted log-likelihood estimator from neighboring observations when no observations are available at the site of interest. However, the weights have to be chosen carefully according to a measure of distance between the distributions. Also, including observations in the log-likelihood that are too distant will hamper the estimators.

2.2.2. *Embedding Space*

We build a space, called the embedding space, so that in that space, two raingauge stations are close in terms of Euclidean distance if the distribution of annual rainfall maxima at the two stations is similar. By defining the distance $d(\cdot,\cdot)$ as the Euclidean distance in the embedding space, we fulfill the requirement that the distance $d(\cdot,\cdot)$ should reflect the distance in distribution, as mentioned in the previous paragraph.

We first define the notion of *similarity in distribution*. We say that two raingauges Z_0 and Z_1 are *similar in distribution* if the Kolmogorov-Smirnov (KS) statistic (see for instance [26]) computed from the observed annual maxima at these raingauges is small. The smallest the statistic is, the more similar the two raingauges are. The KS statistic is the largest vertical distance between the two empirical distribution functions. It thus takes value in $[0,1]$. The KS statistic will only be zero if the two samples are the same which means that even if the two samples are from the same distribution, it is highly unlikely to be equal to zero. In this regard, the KS statistic is not a proper distance.

In order to define coordinates in the embedding space for the sites with observations, we rely on multidimensional scaling (MDS) [5]. MDS algorithms seek to find coordinates in a low dimensional space for n objects starting from a similarity matrix $\mathscr{M}$ between these n objects. If $\mathscr{M}$ is an Euclidean distance matrix then MDS gives the same solution as principal component analysis. In our application of MDS, $\mathscr{M}$ is based on the KS statistic and is not a proper Euclidean distance matrix so we applied a non-metric version of MDS. The non-metric MDS algorithm looks for coordinates which minimize a stress function. One possible stress function is the following [19]:

$$\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathscr{M}_{ij}-d_{ij})^2 \Big/ \sum_{i=1}^{n}\sum_{j=1}^{n}\mathscr{M}_{ij}^2}$$

where $d_{ij} = d(Z_i, Z_j)$ is the Euclidean distance from the coordinates in the MDS space, the space that we use to define the distance in distribution. We use the implementation of the algorithm in the function isoMDS in R [22]. The embedding space dimension is chosen by visually assessing that the shortest distances are well preserved (small KS statistics correspond to small Euclidean

 ISSN: 2102-6238

distances in the embedding space). It is not as important to preserve larger distances. In our applications, the embedding space dimension varies between two and three.

The KS statistic can only be computed for the sites with observations. In order to define embedding coordinates for ungauged sites, we implemented one-layer feed-forward neural networks to learn the mapping from the station geographical coordinates to the MDS embedding coordinates (see [9] for combined application of neural networks and MDS algorithms). These neural networks are a convenient class of functions to learn such a mapping : they are non-parametric non-linear flexible algorithms [3]. Given that the number of hidden units, which controls the complexity of the neural network, is chosen according to the data at hand, the neural network can approximate arbitrarily well any continuous function [16]. For this task, neural networks are trained by minimizing the sum-of-square error between the embedding space coordinates computed with MDS and the estimates provided by the neural network. The number of hidden units of the neural networks are chosen with the cross-validation method [3].

To summarize, here are the steps that we follow to define a distance $d(\cdot,\cdot)$ which can be computed for any site (regardless whether we have observations at that site) :

1. Compute a similarity matrix $\mathscr{M} \in \mathbb{R}^n \times \mathbb{R}^n$ where n is the number of stations where we have observations and $\mathscr{M}_{i,j}$ is the KS statistic between stations Z_i and Z_j.
2. Get the embedding space coordinates for the n stations by applying MDS on the similarity matrix for a given embedding space dimension.
3. Fit a mapping from the station geographical coordinates to the embedding coordinates with a neural network whose number of hidden units is selected via cross-validation.
4. Take an ungauged station Z with known geographical coordinates. Get the embedding coordinates by applying the mapping learned by the neural network.
5. Compute the Euclidean distances on the embedding coordinates with the gauged stations.

The computation load lies in the preprocessing (steps 1-3). Once this is done, the computation of the dissimilarity in distribution for new stations just involve calling the neural network mapping to get the embedding coordinates and computing the Euclidean distances on them. Those last steps are quick.

3. Empirical Study

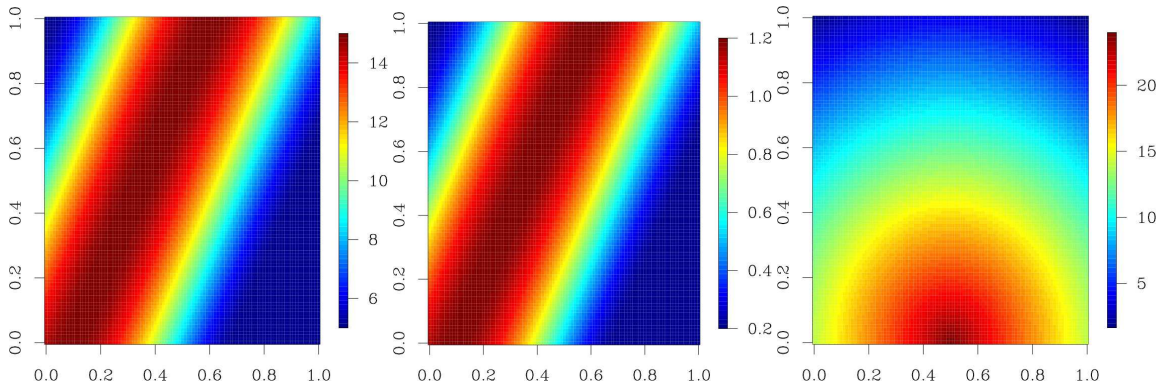
In this section, we study empirically the behavior of the weighted log-likelihood estimator on spatial data. We use the distance $d(\cdot,\cdot)$ defined in the previous section. With a positive shape parameter, the GEV distribution support is $[\theta,+\infty)$ where $\theta = \mu - \sigma/\xi$. When using a kernel such as the Gaussian kernel, all observations have positive weights (which could be negligible for observations from far away stations). This type of kernel forces θ to be smaller than all observations so that they all lie in $[\theta,+\infty)$, otherwise the log-likelihood goes to minus infinity. This means that θ would be almost constant over the whole area. This is not realistic for our application since in the area considered, rainfall intensity is much more important on the mountains than in the river valley. More precisely, we expect θ to be higher in the mountains. We thus resort on the following finite support biquadratic kernel $K(u) = (1-u^2)^2 I_{\{u\leq 1\}}$ which limits the observations with positive weights in the log-likelihood. Since distant observations can

 ISSN: 2102-6238

harm the weighted log-likelihood estimator, (see the one dimensional study 2.2.1), we chose the bandwidth to be as small as possible while ensuring enough observations were kept by the kernel so that the estimation could carry on properly. In practice, we chose the bandwidth to be equal to the standard deviation of the geographical site inter-distances and increase it if not enough weights were positive.

3.1. 2D Synthetic Data Set

We define two synthetic data sets which contain 169 fictitious stations on a 13×13 regularly spaced grid over the unit square $[0,1] \times [0,1]$. The observations at each station follow a GEV distribution whose parameters vary over the unit square. For the first data set A, the location and scale parameters have the same spatial variability although their range of values is different. These functions are represented in Figures 2a and 2b respectively. For the second data set B, we assume a different influence of the space on the location and scale parameters as depicted in Figures 2a and 2c respectively. For these two data sets, the shape parameter is invariant in space and has value $\xi = 0.2$.

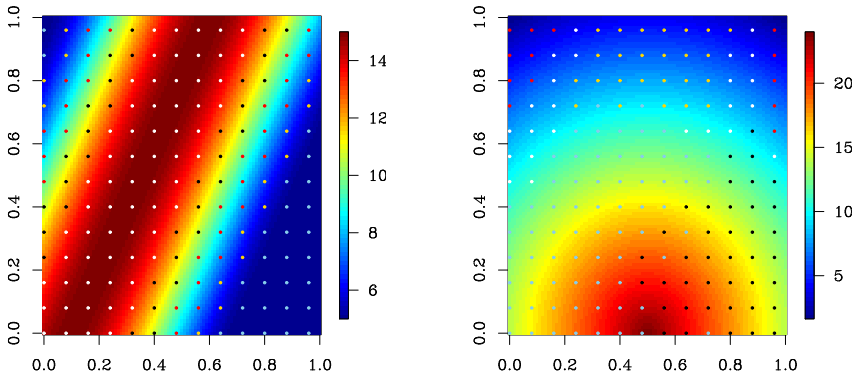


(a) *Location parameter A & B* **(b)** *Scale parameter A* **(c)** *Scale parameter B*

Figure 2: *Spatial mapping of the GEV parameters for the two synthetic data sets. The shape parameter is constant and equal to 0.2, the location parameter takes on the same response surface in both data sets as shown in Figures 2a whereas the scale parameter takes on a different response surface (2b and 2c).*

From a priori investigation of the data, the embedding space dimension is chosen to be two (higher dimension did not improve the results). In order to get an idea of the neighborhood induced by the embedding space, we clustered the 169 fictitious stations into five groups by applying the k-means algorithm (see [15], Chapter 14 for an introduction to k-means) to the station coordinates in the embedding space. The resulting clustering is shown in Figure 3 where the stations with the same color belong to the same cluster. For both synthetic data sets, the clustering reflects the response surfaces of the GEV parameters so that stations with similar GEV parameters tend to belong to the same cluster. For the data set B, the clustering reflects the combined influence of the two distinct spatial pattern of the location and scale parameters.

 ISSN: 2102-6238



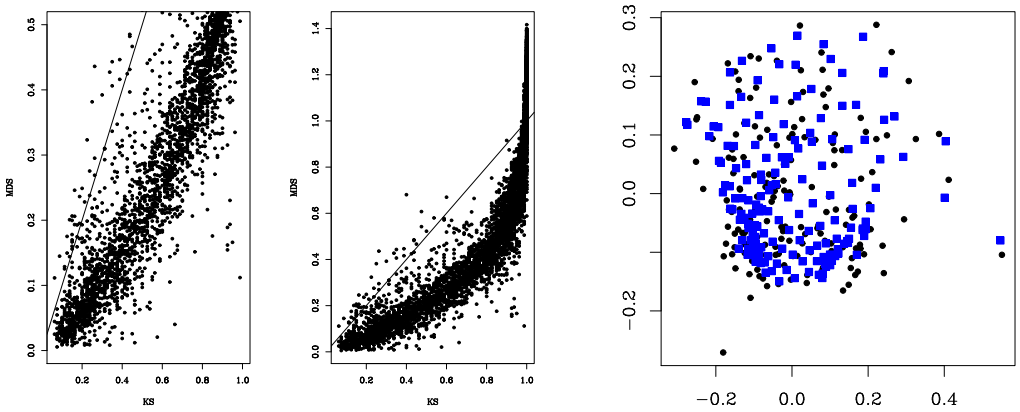
(a) *Data set A* **(b)** *Data set B*

Figure 3: *Clustering of the 169 stations into five groups as indicated by the different colors. The clustering is performed by applying KMeans in the embedding space.*

We apply the leave-one-out method to evaluate the performance of the weighted likelihood approach on the synthetic data sets. For each loop of the leave-one-out method, we performed 10-fold cross-validation to select the number of hidden units of the neural network (it thus results in a double cross-validation method). For one loop of the leave-one-out method, steps 2-3 of the embedding space construction are illustrated in Figure 4 for the synthetic data set B. In Figure 4a, we can check whether the distances in the embedding space match the KS statistics. Small KS statistics should correspond to small embedding distances. The stations mapped into the embedding space by the MDS algorithm are represented as black dots in Figure 4b. A neural network approximates the mapping from the station coordinates in $[0,1] \times [0,1]$ to the embedding coordinates in $\mathbb{R}^2$. Eight hidden units were selected by the cross-validation method. The embedding coordinates of the stations learned by a neural network are shown as blue squares in Figure 4b.

The GEV parameter surface responses learned through the leave-one-out method are shown in Figure 5 and Figure 6. Globally, the surface responses are fairly well reproduced even for the more challenging data set B although the shape parameter is more tricky to estimate. As a measure of performance, we computed the average difference in log-likelihood between the generative model and the learned model. These are displayed in Figure 7. Values around zero indicate that the learned model is equivalent to the generative model in terms of log-likelihood. Finally, we compare the GEV parameter estimates obtained from the weighted log-likelihood method with the estimates obtained by maximizing the log-likelihood on the local observations. The comparisons for the synthetic data set A are shown in Figure 8 (comparisons for data set B give similar results). The weighted log-likelihood estimates are less spread out than their local counterparts, the most striking example being the shape parameter estimate.

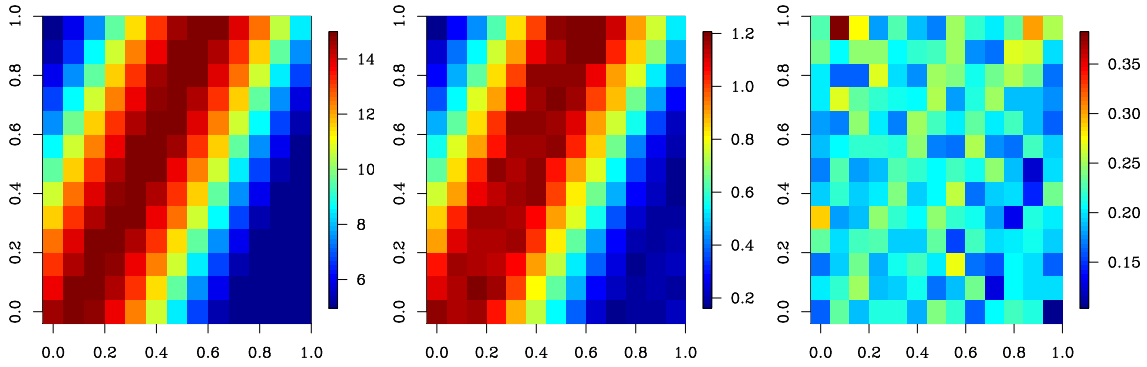
 ISSN: 2102-6238



(a) *Comparison of the KS statistics (x-axis) versus the Euclidean distances in the embedding space (y-axis). Left panel shows a zoom on the smallest values which should be preserved as well as possible.*

(b) *Embedding space : black dots represent the embedding from the MDS algorithm and the blue squares the corresponding estimates from a neural network with 8 hidden units.*

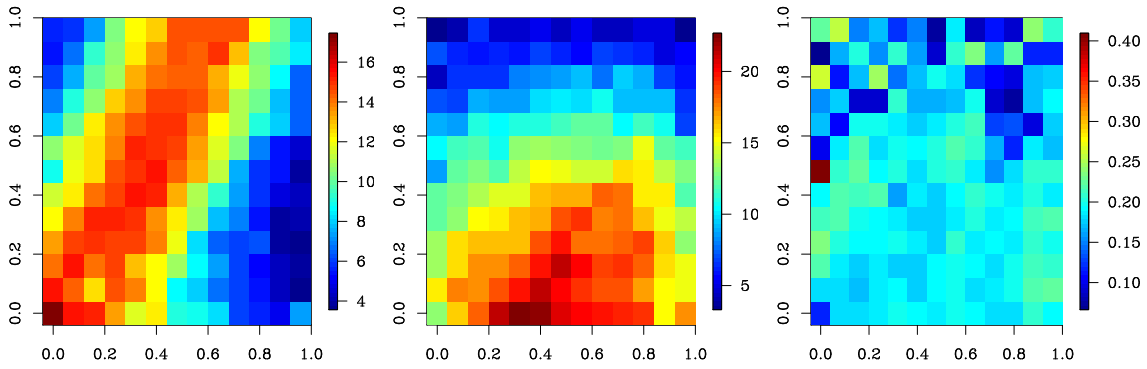
Figure 4: *Embedding space computation with the Multidimensional scaling (MDS) algorithm applied to the Kolmogorov-Smirnov (KS) statistics as a measure of similarity on the synthetic data set B. The embedding space has dimension 2.*



(a) *Location parameter* **(b)** *Scale parameter* **(c)** *Shape parameter*

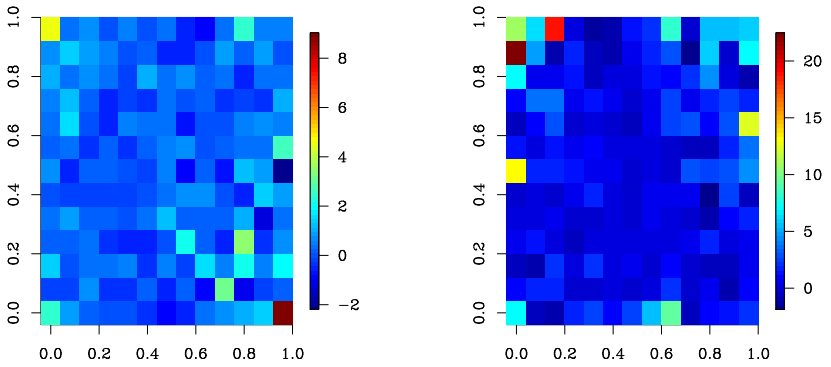
Figure 5: *Leave-one-out estimation of the GEV parameters on the* 13×13 *fictitious stations of the synthetic data set A with the weighted log-likelihood estimator.*

 ISSN: 2102-6238



(a) *Location parameter* **(b)** *Scale parameter* **(c)** *Shape parameter*

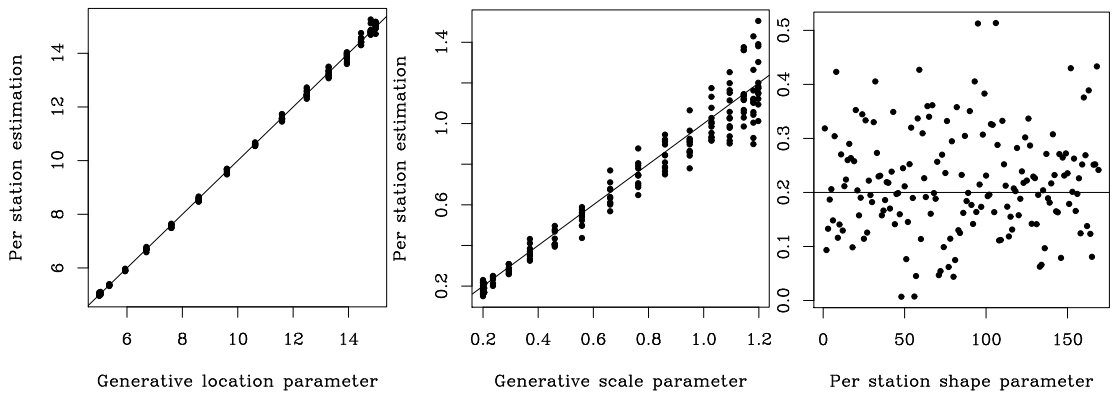
Figure 6: *Leave-one-out estimation of the GEV parameters on the* 13×13 *fictitious stations of the synthetic data set B with the weighted log-likelihood estimator.*



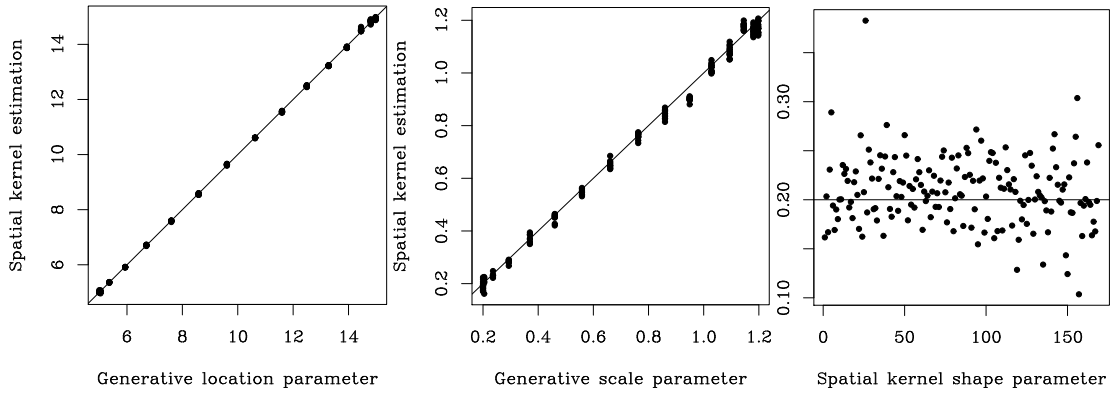
(a) *Data set A* **(b)** *Data set B*

Figure 7: *Average difference in log-likelihood between the generative model and the learned model. Values close to zero means that the learned model is equivalent to the generative model in terms of log-likelihood.*

 ISSN: 2102-6238



(a) *Unconditional log-likelihood estimates*



(b) *Weighted log-likelihood estimates*

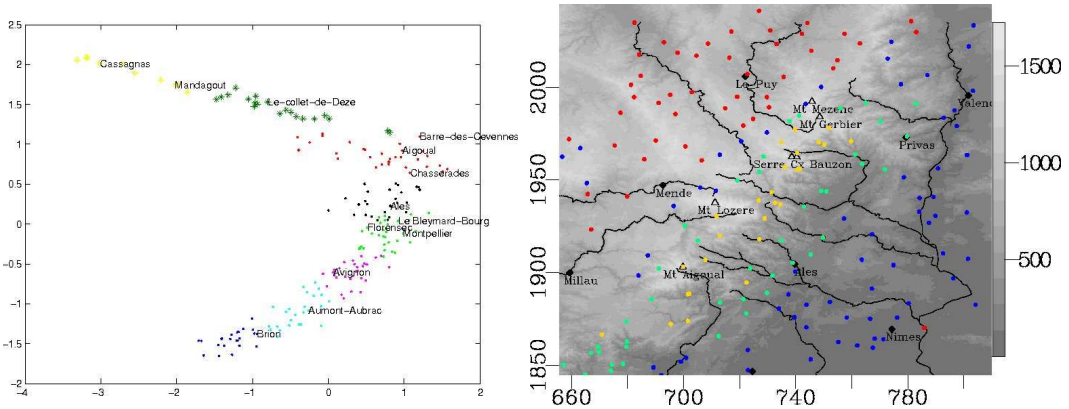
Figure 8: *Comparison of GEV parameter estimates versus parameters of the generative model. Location and scale parameters are compared by means of QQ-plots (column one and two respectively) whereas the shape parameter estimates are compared to the constant shape parameter of the generative model, as shown by the horizontal line in the third column. In the upper panel, the estimates are obtained by maximizing the log-likelihood over the observations at the station (local or unconditional estimation). In the lower panel, the estimates are given by the weighted log-likelihood method that is, local observations are considered missing and only observations from neighboring sites contribute.*

3.2. Rainfall Application

Extreme rainfalls are prevalent in the Cévennes-Vivarais area in the French Mediterranean region. In a predominantly dry climate, sudden and violent rainfalls occur at the foothills and in the Rhône river valley brought by Mediterranean winds loaded with humidity. We have daily rainfall data at 198 stations recorded from 1948 to 1991. The geographical coordinates of the stations are given in Lambert II extended and range from 656 km to 810 km in longitude and from 1845 km to 2045 km in latitude. The stations are represented in the embedding space in Figure 9a. Stations in that space are close to one another if they share similar annual maxima distributions. In Figure 9b,

 ISSN: 2102-6238

a four-class clustering illustrates the concept of neighborhood as defined by the embedding space. The clusters are spatially coherent with regards to the geographical space although no direct geographical information is used to build the embedding space. The geographical information is obtained indirectly through the distribution of the annual maxima. The cluster in yellow covers the mountain summits of the Cévennes, the green one is at the foothills, the blue cluster spreads in the Rhône river valley and the red one lays in the back country.

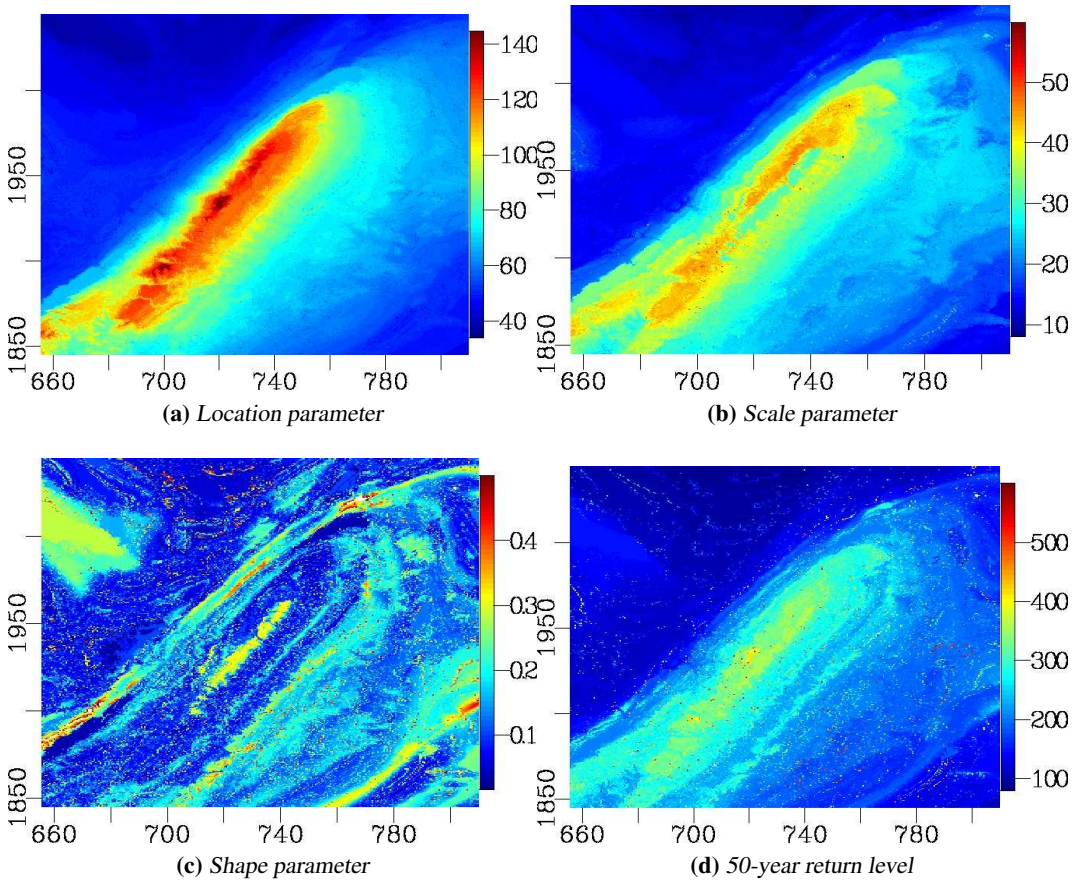


(a) *Embedding Space* **(b)** *Clustering*

Figure 9: *Cévennes-Vivarais annual maxima rainfall data : embedding space according to the KS statistics as a measure of similarity (Figure 9a) and four-class clustering on the embedding coordinates represented in the geographical space (Figure 9b).*

We learned the response surface of the GEV parameters with the weighted likelihood estimator over a 294×381 grid which spans the area covered by the stations. The embedding space dimension was chosen to be three from visual investigation of the match between the smallest KS statistics and their corresponding embedding space distances. The mapping from the station coordinates *longitude, latitude* and *altitude* to the embedding space coordinates was learned with a neural network. Eight hidden units were chosen with the 10-fold cross-validation method. The learned location, scale and shape parameter surfaces are shown in Figures 10a-10c respectively. The location parameter (Figure 10a) is strongly influenced by the topography : the highest values correspond to the highest altitudes. The scale parameter (Figure 10b) displays a similar pattern albeit with some variation at the mountain summit showing a mild depression. The surface response of the shape parameter is more contrasted but the shape of the mountain range is still clearly visible. Finally, we have computed the 50-year return level in Figure 10d which combines the information from all three GEV parameters. Here again, we see the strong influence of the mountain range in determining the value of the 50-year return level. These results are consistent with expert knowledge in the area : the Cévennes mountain range acts as a barrier that blocks the humidity coming from the Mediterranean sea and triggers precipitation.

 ISSN: 2102-6238



(a) *Location parameter* **(b)** *Scale parameter*

(c) *Shape parameter* **(d)** *50-year return level*

Figure 10: *Cévennes-Vivarais GEV parameter surface responses and 50-year return level determined by the weighted log-likelihood estimator.*

4. Conclusion

We have presented a new estimator of spatial extreme quantiles involving both nonparametric statistical techniques (kernel weighted log-likelihood) and machine-learning methods (MDS, neural networks). Further work would however be needed to obtain a fully data-driven estimator. In particular, the automatic choice of the bandwidth h is an open issue. The optimal bandwidth should achieve a tradeoff between the bias and the variance of the estimator. We also plan to improve the construction of the embedding coordinates by using nonlinear methods such as Locally Linear Embedding (LLE) [23] or t-Distributed Stochastic Neighbor Embedding (t-SNE) [28] instead of MDS.

Finally, the results on the Cévennes-Vivarais annual maxima rainfall data may be improved by the introduction of more covariates describing the topography or the atmospheric conditions. It may be also of interest to introduce regularity constraints on the response surfaces. For instance, in such rainfall applications, the parameter ξ is often supposed to be constant on the region of interest.

 ISSN: 2102-6238

References

[1] J. Beirlant and Y. Goegebeur. Regression with response distributions of Pareto-type. *Computational Statistics and Data Analysis*, 42:595–619, 2003.

[2] J. Beirlant and Y. Goegebeur. Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis*, 89:97–118, 2004.

[3] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford, 1995.

[4] V. Chavez-Demoulin and A.C. A.C. Davison. Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, series C*, 54:207–222, 2005.

[5] C. Chen, W. Härdle, A. Unwin, M. A. A. Cox, and T. F. Cox. Multidimensional scaling. In *Handbook of Data Visualization*, Springer Handbooks of Computational Statistics, pages 315–347. Springer Berlin Heidelberg, 2008.

[6] A. Daouia, L. Gardes, S. Girard, and A. Lekina. Kernel estimators of extreme level curves. *Test*, 20(14):311–333, 2011.

[7] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B*, 52(3):393–442, 1990.

[8] A.C. Davison and N.I. Ramesh. Local likelihood smoothing of sample extremes. *Journal of the Royal Statistical Society, series B*, 62:191–208, 2000.

[9] D. de Ridder and R.P.W. Duin. Sammon's mapping using neural networks: A comparison. *Pattern Recognition Letters*, 18:1307–1316, 1997.

[10] A. Dekkers and L. de Haan. On the estimation of the extreme-value index and large quantile estimation. *The Annals of Statistics*, 17(4):1795–1832, 1989.

[11] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*. R.E. Krieger publishing company, 1987.

[12] L. Gardes and S. Girard. Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. *Extremes*, 13(2):177–204, 2010.

[13] M. Guida and M. Longo. Estimation of probability tails based on generalized extreme value distributions. *Reliability Engineering and System Safety*, 20:219–242, 1988.

[14] P. Hall and N. Tajvidi. Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, 15:153–167, 2000.

[15] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27:83–85, 2005.

[16] K. M. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[17] J.R.M. Hosking. Algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics*, 34:301–310, 1985.

[18] F. Hu and J. V. Zidek. The weighted likelihood. *The Canadian Journal of Statistics*, 30(3):347–361, 2002.

[19] J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to nonmetric hypothesis. *Psychometrika*, 29:1–27,115–129, 1964.

[20] A.J. Macleod. AS R76– a remark on algorithm AS 215: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics*, 38:198–199, 1989.

[21] P. Prescott and A.T. Walden. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, 67:723–724, 1980.

[22] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[23] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[24] R. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). *Statistical Science*, 4:367–393, 1989.

[25] J. Staniswalis. The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association*, 84(405):276–283, 1989.

 ISSN: 2102-6238

[26] M.A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974.

[27] Y. Tramblay, L. Neppel, and J. Carreau. Climatic covariates for the frequency analysis of heavy rainfall events in the mediterranean region. *Natural Hazards and Earth System Sciences*, 2011.

[28] L.J.P. van der Maaten and G.E. Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[29] I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364):812–815, 1978.

[30] G. Wotling, Ch. Bouvier, J. Danloux, and J.-M. Fritsch. Regionalization of extreme precipitation distribution using the principal components of topographical environment. *Journal of Hydrology*, 233:86–101, 2000.

[31] C. Zhou. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 100:794–815, 2009.

[32] C. Zhou. The extent of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis*, 101:971–983, 2010.

Acknowledgment

This work was partially supported by the Agence Nationale de la Recherche (French Research Agency) through its VMC program (Vulnérabilité, Milieu, Climat).

 ISSN: 2102-6238